DOCUMENT RESUME

ED 160 656 TM 007 946

AUTHOR Davis, B. Glen; Trimble, C. Scott
TITLE Consumable Booklets vs Single Answer Sheets with Non-Consumable Booklets. (Research Report).
INSTITUTION Kentucky State Dept. of Education, Frankfort. Div. of Evaluation.
PUB DATE Jul 78
NOTE 27p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Achievement Tests; *Answer Sheets; Cost Effectiveness; Grade 3; Primary Education; Raw Scores; *Response Style (Tests); *Student Ability; *Testing; *Tests; Verbal Ability
IDENTIFIERS Comprehensive Tests of Basic Skills

ABSTRACT

The Comprehensive Test of Basic Skills (CTBS) was administered to over 200 third graders to investigate the effects of using consumable test booklets versus non-consumable booklets which require a separate answer sheet. It was hypothesized that there would be no significant differences in CTBS scores between students using consumable booklets and those using single answer sheets. Raw scores on five verbal subtests were observed, as well as the number of responses left blank (indicating students who worked at a slower rate). Analyses were conducted for six groups of students: (1) entire sample; (2) low verbal ability; (3) medium ability; (4) high ability; (5) educable mentally handicapped, learning disabled, and those who had been retained from grade promotion; and (6) entire sample except group (5). Although raw scores of the entire group were slightly higher when consumable booklets were used, the mean difference was not significant. More items were left blank by students using answer sheets, particularly for the low and medium ability groups. For students with learning difficulties, scores were significantly higher on the spelling test only when consumable booklets were used. It was concluded that there was no evidence to justify the use of one format over the other at the third grade level. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160656

CONSUMABLE BOOKLETS

vs

SINGLE ANSWER SHEETS WITH NON-CONSUMABLE BOOKLETS

(Research Report)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
E. Norman Sims
Ky Dept of Ed
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Prepared by

B. Glen Davis
C. Scott Trimble

Research Unit

Office of Research and Planning
Donald Van Fleet, Head

July 1978

TM007 946

KENTUCKY DEPARTMENT OF EDUCATION
James B. Graham, Superintendent

TABLE OF CONTENTS

Page

BACKGROUND 1

RATIONALE 2

PROBLEM STATEMENT 3

HYPOTHESIS 3

LITERATURE REVIEW 4

METHODOLOGY AND DESIGN 5

RESULTS 7

TABLE 1A - Comparison of Mean Raw Scores of All Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets . . 8

TABLE 1B - Comparison of Mean Number of Blank Responses of All Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets 10

TABLE 2A - Comparison of Mean Raw Scores of All Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets Except EMH's, LD's, and Retentions 12

TABLE 2B - Comparison of the Mean Number of Blank Responses for Third Graders Using Consumable Booklets and Separate Answer Sheets Except EMH's, LD's, and Retentions 12

TABLE 3A - Comparison of Mean Raw Scores of Low Ability Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets 13

TABLE 3B - Comparison of Mean Number of Blank Responses for Low Ability Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets 13

TABLE 4A - Comparison of Mean Raw Scores of Medium Ability Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets 14

TABLE 4B - Comparison of Mean Number of Blank Responses of Medium Ability Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets 14

TABLE 5A - Comparison of Mean Raw Scores of High Ability Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets 15

TABLE 5B - Comparison of Mean Blank Responses of High Ability Third Graders Using Consumable Booklets and Those Using Separate Answer Sheets. 15

TABLE 6A - Comparison of Mean Raw Scores of Third Grade EMH's, LD's, and Retentions Using Consumable Booklets and Those Using Separate Answer Sheets. 18

TABLE 6B - Comparison of Mean Number of Blank Responses of Third Grade EMH's, LD's, and Retentions Using Consumable Booklets and Those Using Separate Answer Sheets 18

TABLE 7 - Comparison of the Upper Limits of Third Grade Raw Score Deciles Between Students Using Consumable Booklets and Those Using Separate Answer Sheets. 19

TEACHER SURVEY 20

SUMMARY AND CONCLUSIONS. 21

REFERENCES 23

BACKGROUND

Pilot testing programs were begun in the fall of 1966, under Wayne Ashley. Twenty-one pilot school districts participated in this program in October of 1966, with the State Department of Education furnishing all materials. The California Short Form of Mental Maturity and the California Achievement Test were administered during this pilot study. Scoring of tests was contracted to an out-of-state concern.

This program was begun as a result of a recommendation of the Cresap, McCormick and Paget report on quality education submitted to the Department of Education in the summer of 1966.

In the fall of 1967, Mr. Charles Morgan was hired by the State Department of Education as administrator for the pilot testing program. Thirty-one local districts participated in the program during the fall of 1967. The success of this pilot program resulted in the opening of this test scoring service to all districts that wanted to participate on a volunteer basis, during the fall of 1968. Local educational agencies purchased their own materials at this point. During that particular school year, fifty-one districts participated. Free scoring was made available on the fourth, eighth, and eleventh grade levels. This program became known as the Statewide Volunteer Testing Program.

In 1969, the Department of Education introduced, in addition to the California Achievement Test, the Comprehensive Test of Basic Skills (CTBS), the Short Form Test of Academic Aptitude (SFTAA), the Otis Lenon IQ and the Stanford Achievement Test (64th edition).

In 1970, the Department of Education attempted to score all these tests, and the volume was overwhelming. More school districts participated this year than ever before. This resulted in the Department making a decision to score only the SFTAA and CTBS in Grades 4, 8, and 11 during the 1971-72 school year.

In 1973, a pilot program was undertaken in Grades 1, 2, and 3 for scoring the CTBS lower levels. This research endeavor was successful. Due to this, in 1974, the Department scored tests from students in Grades 1, 2, and 3 that were participating in Title I programs. In 1975, the Department started scoring Grades 1, 2, and 3 on a volunteer basis for all students. Grades 1, 2, and 3 require special arrangements and procedures for scoring that are different from those utilized in scoring tests for Grades 4 and above. These three grades used a consumable booklet, which when separated, became 12 documents for each student, as compared to one document per student in Grades 4 and above. Free scoring of the lower levels continued up through the 1977-78 school term. On this volunteer basis, these three grades resulted in scoring approximately 900,000 documents per year.

The Kentucky State Legislature during its regular biennial session in 1978 passed legislation entitled the Educational Improvement Act. This legislation requires that the State Department of Education test all students in Kentucky Public Schools which are enrolled in Grades 3, 5, 7, and 10 beginning during the 1978-79 school term.

RATIONALE

The State Legislature, having mandated the testing of students at Grades 3, 5, 7, and 10, left the Department of Education no alternative but to prepare to score all these instruments on a statewide basis. In the past, third grade students have utilized a consumable booklet in taking tests, which is separated, and results in 12 separate documents to be scored for each student. This in itself results in considerable time and expense for the Testing Unit. If tests given to this grade could be scored on a single answer sheet as Grades 5, 7, and 10 can be, this would reduce processing time and expense considerably, not only

in the scoring, but in the purchasing of reusable test booklets instead of consumable ones. Membership at the close of the first month during the 1977-78 school term reveals that there are 52,084 third graders in Kentucky Public Schools. Using this information only, one could see that if the single answer sheet was utilized, there would be 52,084 documents scored. However, if the Department is required to utilize the present consumable booklet, which requires 12 separate documents per individual, this would result in over 625,000 documents necessary to be scored for the third grade only.

By using the 1977-78 fall membership of Kentucky Public Schools only, the total number of students in Grades 3, 5, 7, and 10 for the 1978-79 school term would be approximately 223,318. If the third grade students could utilize the test requiring only one answer document, this would reduce the work load from 834,126 documents to 223, 318 or, in other words, a document load reduction of 73 percent for the grades mandated to be tested.

PROBLEM STATEMENT

As a result of the rationale above, the research problem resulting herein was stated in the following manner:

> Is there a significant difference between the scores of third grade students utilizing CTBS Form S Level 1 consumable booklets and those utilizing CTBS Form S Level 1 non-consumable booklets which require a separate answer sheet?

HYPOTHESIS (Stated in the null form)

There are no significant differences between third grade students' CTBS scores utilizing the consumable booklets and those using single answer sheets with the non-consumable booklets.

LITERATURE REVIEW

An exhaustive search of the literature was made utilizing the Educational Research Information Center (ERIC), Research and Information Services for Education (RISE), and Current Index to Journals in Education (CIJE). There was an almost total void in literature concerning research conducted which addressed itself to the particular research problem stated above. While it was determined that at least one state in our area used the non-consumable booklets which required a separate answer sheet for testing third grade students, there was no hard data available to posit the effectiveness of this method as compared to the use of consumable booklets.

While there was apparently no published research results concerning this specific problem in the information sources reviewed, some relevant data was found elsewhere. Research was conducted using the Iowa Tests of Basic Skills to determine the ability of third grade pupils to use separate answer sheets (Hieronymus). The research groups in this case were equated with respect to average achievement in reading. There were no significant differences between the means in this study. However, there were considerably more omissions for the answer sheet procedure which is to be expected because recording answers on separate answer sheets requires more time. One might reasonably assume that the answer sheet procedure could result in less guessing.

When grade equivalent scores were used as the criterion measure to determine the possibility of third grade students using a separate answer sheet as compared to a consumable booklet, pupils who marked responses directly on the test booklet tended to make higher grade equivalent scores (Aleyideino, 1968). Statistically significant differences were observed in the case of subtests which were comprised of more time-consuming items.

The lack of corroborating data and the scarcity of published results, left us no alternative but to design a research endeavor which would support any

decision made by the Department of Education concerning this particular problem.

METHODOLOGY AND DESIGN

One school district in the Commonwealth of Kentucky which was not involved in spring testing, was selected for the purpose of conducting this research endeavor. It was necessary to select a district which had a third grade enrollment of at least 200 for the purpose of stratifying the sample. The sample, in this case, included all third grade students in the particular school district. The sample was stratified by verbal IQ scores, those scoring one standard deviation (SD) or more above the mean, those scoring within the range of one S.D. above or below the mean, and those scoring one S.D. or more below the mean. These IQ scores were determined by administering the SFTAA Level 2, hand-scored IQ tests in all third grade classrooms within the particular system selected. The administering of the IQ tests was necessary to arrive at the sample to be tested with the CTBS using a matched-group technique. Only the language portion of the IQ was used in arriving at these matched groups.

This particular process, scoring and matching of groups, was completed in two weeks after the IQ test was given. After the matched groups had been derived, it was necessary to go back into the same school system and administer the CTBS Form S to all third grade classrooms, with one group of students utilizing the consumable booklets and the other utilizing the non-consumable booklets which require the use of a separate answer sheet.

Information was also gathered from the classroom teachers (who administered the tests) so as to be able to identify those students who were classified as educationally and mentally retarded (EMH) or as having learning disabilities (LD) and to be able to identify those students who had been retained at any grade. Although the major interest is in those differences between third graders taking

the consumable booklet and those taking the separate answer sheet, the study is also interested in differences which might appear in various subgroups within the third grade. For this reason, the following analyses were undertaken:

(1) an analysis including the entire third grade;

(2) an analysis including those third graders who were not EMH or LD and had not been retained;

(3) an analysis including only low ability students;

(4) an analysis including only medium ability students;

(5) an analysis including only high ability students;

(6) an analysis of only third graders classified as EMH, LD, and/or retentions.

The methodology used to determine if a statistically significant difference exists between students using the consumable booklet and those using the separate answer sheet was the independent t test. These calculations were completed through the use of SAS's PROCEDURE T TEST, (Barr, 1976). The variance of the two groups was assumed to be equal. Before a difference was considered to be statistically significant, the probability of obtaining an associated t value by chance had to be less than .05.

Deciles for both groups, consumable booklet and separate answer sheet were also calculated. This was done to further assist in looking for any possible differences in the distribution of the two sets of scores. Again, SAS was used to perform the necessary calculations.

RESULTS

The *t* test for independent measures was applied to the previously defined subdivisions of the third grade sample to determine if there was any statistically significant differences between the mean scores of students using consumable booklets and those using separate answer sheets on the five subtests of CTBS Form S Level 1. Only five subtests measuring student achievement in the basic skills have been observed for purposes of this study. These subtests were: (1) reading vocabulary, (2) reading comprehension, (3) spelling, (4) language mechanics, and (5) language expression. The analysis was limited to these five subtests because student ability had been defined according to student scores on the verbal portion of the Short Form Test of Academic Aptitude. Students were considred to be of "low ability" if their verbal IQ's fell below 84, of "medium ability" if their verbal IQ's fell between 84 and 116, and of "high ability" if their verbal IQ's fell above 116.

The first analysis undertaken included <u>all</u> third graders. Tables 1A and 1B include the findings of the analysis. Table 1A reports those differences observed between students using consumable booklets and students using separate answer sheets. As can be observed, there are no statistically significant differences reported in Table 1A. In all subtests except language mechanics, students using consumable booklets scored slightly higher than those using separate answer sheets. In language mechanics, the students using the separate answer sheets scored slightly higher. In no case was the mean raw score difference significant.

These results are not consistent with the Aleyideino study, (1968). Aleyideino looked for differences among third grade students using the consumable booklet format and those using the separate answer sheet format of the Iowa Tests of Basic Skills. Significant differences were found in nine of the eleven subtests considered.

TABLE 1A: COMPARISON OF MEAN RAW SCORES OF ALL THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Raw Score	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	138	25.01	8.65	1.16	1.05
Answer Sheet	136	23.85	9.47		
Reading Comprehension					
Consumable Bklt.	138	27.84	10.87	.69	.52
Answer Sheet	136	27.15	11.39		
Total Reading					
Consumable Bklt.	138	52.85	18.93	1.85	.78
Answer Sheet	136	51.00	20.13		
Spelling					
Consumable Bklt.	137	33.18	7.19	1.86	1.94
Answer Sheet	136	31.32	8.66		
Language Mechanics					
Consumable Bklt.	138	12.36	5.03	-.27	-.45
Answer Sheet	134	12.63	4.91		
Language Expression					
Consumable Bklt.	139	20.46	6.41	.81	1.03
Answer Sheet	134	19.65	6.55		
Total Language					
Consumable Bklt.	135	66.23	16.36	2.66	1.26
Answer Sheet	132	63.57	18.14		

* Asterisk indicates that mean difference is significantly different from zero at the .01 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

In all nine cases, students using consumable booklets scored higher than those using the separate answer sheet. No significant differences were found in the spelling and punctuation subtest scores.

Three important differences between Aleyideino's study and this one must be stressed. First, Aleyideino's sample was considerably larger, over twice as large. Second, Aleyideino reports that he used grade equivalent scores in the analysis while raw scores were used in this study. Third, and most important, Aleyideino used a one-tailed t test to determine significance, while a two-tailed t test was used in this study. The decision as to which test is appropriate depends on how the null hypothesis is stated in mathematical terms. If the null hypothesis is stated as:

H_o: Mean consumable booklet score minus mean answer sheet is less than or equal to zero

then, a one-tailed t test is appropriate. In this case, one is not concerned

with the possibility of observing differences which could occur as a result of students using the separate answer sheet scoring higher than students using consumable booklets. Such differences would show up as a negative difference and as the above null hypothesis is stated would be considered supportive evidence.

Aleyideino apparently stated the null hypothesis in a similar format. As was done in this study, the null hypothesis can be stated as:

H_o: Mean consumable booklet score minus mean separate answer sheet score equals zero

When stated in this manner, a two-tailed *t* test is appropriate. In this study, the emphasis was on observing any significant mean difference which might have occurred in either direction. Thus, equal emphasis is placed on the possiblity that consumable booklet scores are higher than separate answer sheet scores and on the reverse pattern.

Table 1B summarizes the mean number of blank responses left by the students using the two test formats. Number of blank responses is considered to be an indicator of work rate. Students leaving more blanks are assumed to have worked slower. In the reading vocabulary and spelling subtests, students using the answer sheet format left a significantly higher number of blank responses than did those using the consumable booklets. In the remaining three subtests, the observed differences were insignificant and could be attributed to chance. These findings are also somewhat inconsistent with the Aleyideino study. He found that in nine of the eleven subtests of the Iowa Tests of Basic Skills, the full test work rates (average number of items per minute) differed significantly. Only on the arithmetic computation did students using separate answer sheets work significantly faster than those using consumable booklets. In the remaining eight instances where Aleyideino observed significant differences, students using consumable booklets had higher rates of work. The inconsistence between this study and the Aleyideino study lies in the degree to

which significant differences were observed. Aleyideino found significant differences in work rates on nine of the eleven subtests considered. In this study, significant differences were observed in only two of the five subtests considered.

TABLE 1B: COMPARISON OF MEAN NUMBER OF BLANK RESPONSES OF ALL THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Blank Responses	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	138	.78	2.31	-1.10	-2.58 *
Answer Sheet	136	1.88	4.47		
Reading Comprehension					
Consumable Bklt.	138	1.14	2.68	-.21	-.52
Answer Sheet	136	1.35	3.81		
Spelling					
Consumable Bklt.	137	.60	1.60	-2.00	-4.16 *
Answer Sheet	136	2.60	5.38		
Language Mechanics					
Consumable Bklt.	138	.23	.72	.05	.51
Answer Sheet	134	.18	.96		
Language Expression					
Consumable Bklt.	139	.51	1.38	.17	1.22
Answer Sheet	134	.34	.94		

* Asterisk indicates that mean difference is significantly different from zero at the .05 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

The above-mentioned analysis included <u>all</u> third graders: students who are considered educationally and mentally handicapped (EMH), students having learning disabilities (LD), students who had been retained one or more times, and students who had progressed through the first three grades normally. This analysis was undertaken becasue at the time of this study no decision had been made as to whether or not certain types of students would be exempt from the statewide testing mandated by the Educational Improvement Act. In the absence of this decision,

a second analysis was undertaken including only those "normal" third graders. Normal refers to those students who are not classified as EMH or LD, and who had not been retained at any grade level so far. This was done to give state decision-makers some idea about what kinds of differences could be expected between student scores on the two different test formats if students identified as having some kinds of learning problems were exempted from the testing program.

Tables 2A and 2B include a summary of the analysis which excluded all EMH's, LD's, and retentions. The pattern of statistically significant differences did not change from that observed in the first analysis. No significant differences between mean raw scores were observed on the five subtests and two CTBS total scores. The same pattern of significant differences in mean number of blank responses was again observed. Students using separate answer sheets left a significantly higher number of blank responses on the reading vocabulary and spelling subtests than did those using consumable booklets.

A third analysis was undertaken which included only those students considered to be low ability students. Low ability students were defined as those obtaining a verbal IQ score of less than 84 on the Short Form Test of Academic Aptitude (SFTAA). This analysis was done because the emphasis of the Educational Improvement Act which mandates the testing program is on the identification of educational needs. Presumably most educational needs will be identified among students of low ability. Thus, it is most important to observe the performance of low ability students on the two test formats.

Interest in possible differences in the performance levels of low ability students helps generate interest in potential differences among students of other ability levels. For this reason, fourth and fifth analyses were also completed. The fourth analysis included only the students of medium ability as indicated by a verbal IQ falling between 84 and 116 on the SFTAA. The fifth analysis

TABLE 2A: COMPARISON OF MEAN RAW SCORES OF ALL THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS EXCEPT EMH's, LD's, AND RETENTIONS

Subtest	N	Mean Raw Score	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	114	26.38	8.11	.91	.80
Answer Sheet	114	25.47	8.86		
Reading Comprehension					
Consumable Bklt.	114	29.80	10.10	.75	.54
Answer Sheet	114	29.05	10.67		
Total Reading					
Consumable Bklt.	114	56.18	17.56	1.65	.69
Answer Sheet	114	54.53	18.71		
Spelling					
Consumable Bklt.	114	33.97	6.96	1.34	1.33
Answer Sheet	114	32.63	8.21		
Language Mechanics					
Consumable Bklt.	115	13.09	4.69	-.49	-.82
Answer Sheet	112	13.58	4.39		
Language Expression					
Consumable Bklt.	116	21.59	5.81	.70	.89
Answer Sheet	112	20.89	6.02		
Total Language					
Consumable Bklt.	113	68.75	15.27	1.59	.74
Answer Sheet	110	67.16	16.60		

TABLE 2B: COMPARISON OF THE MEAN NUMBER OF BLANK RESPONSES FOR THIRD GRADERS USING CONSUMABLE BOOKLETS AND SEPARATE ANSWER SHEETS EXCEPT EMH's, LD's, AND RETENTIONS

Subtest	N	Mean Blank Responses	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	114	.61	2.00	-.94	-2.29 *
Answer Sheet	114	1.55	3.89		
Reading Comprehension					
Consumable Bklt.	114	.72	1.75	-.46	-1.23
Answer Sheet	114	1.18	3.63		
Spelling					
Consumable Bklt.	114	.46	1.44	-1.97	-3.91 *
Answer Sheet	114	2.43	5.17		
Language Mechanics					
Consumable Bklt.	115	.18	.64	-.03	-.20
Answer Sheet	112	.21	1.05		
Language Expression					
Consumable Bklt.	116	.26	.56	-.03	-.28
Answer Sheet	112	.29	.88		

* Asterisk indicates that mean difference is significantly different from zero at the .05 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

included those students of high ability as indicated by a verbal IQ score greater than 116 on the SFTAA. The third, fourth, and fifth analyses are summarized in Tables 3A and 3B, Tables 4A and 4B, and Tables 5A and 5B respectively.

TABLE 3A: COMPARISON OF MEAN RAW SCORES OF LOW ABILITY THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Raw Score	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	14	11.07	4.32	.64	.45
Answer Sheet	21	10.43	4.08		
Reading Comprehension					
Consumable Bklt.	14	11.00	3.78	-1.14	-.79
Answer Sheet	21	12.14	4.43		
Total Reading					
Consumable Bklt.	14	22.07	5.23	-.50	-.24
Answer Sheet	21	22.57	6.70		
Spelling					
Consumable Bklt.	13	24.38	5.56	3.05	1.30
Answer Sheet	21	21.33	7.22		
Language Mechanics					
Consumable Bklt.	14	5.86	3.11	-.39	-.45
Answer Sheet	20	6.25	2.02		
Language Expression					
Consumable Bklt.	14	8.50	2.38	-1.40	-1.42
Answer Sheet	20	9.90	3.11		
Total Language					
Consumable Bklt.	13	38.62	7.19	1.02	.32
Answer Sheet	20	37.60	9.99		

TABLE 3B: COMPARISON OF MEAN NUMBER OF BLANK RESPONSES FOR LOW ABILITY THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Blank Responses	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	14	1.79	4.35	-1.88	-.98
Answer Sheet	21	3.67	6.26		
Reading Comprehension					
Consumable Bklt.	14	4.14	5.49	3.28	2.58 *
Answer Sheet	21	.86	1.71		
Spelling					
Consumable Bklt.	13	1.62	2.69	-4.05	-1.67
Answer Sheet	21	5.67	8.43		
Language Mechanics					
Consumable Bklt.	14	.64	1.22	.49	1.72
Answer Sheet	20	.15	.37		
Language Expression					
Consumable Bklt.	14	2.50	3.41	1.95	2.31
Answer Sheet	20	.55	1.39		

* Asterisk indicates that mean difference is significantly different from zero at the .01 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

TABLE 4A: COMPARISON OF MEAN RAW SCORES OF MEDIUM ABILITY THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Raw Score	S.D.	Mean Difference	*t*
Reading Vocabulary					
Consumable Bklt.	96	24.22	6.74	.40	.39
Answer Sheet	88	23.82	7.27		
Reading Comprehension					
Consumable Bklt.	96	26.81	8.93	-.25	-.18
Answer Sheet	88	27.06	9.63		
Total Reading					
Consumable Bklt.	96	51.03	14.91	.15	.07
Answer Sheet	88	50.88	15.85		
Spelling					
Consumable Bklt.	96	32.15	6.24	.90	.89
Answer Sheet	88	31.25	7.44		
Language Mechanics					
Consumable Bklt.	97	11.88	4.46	-.85	-1.32
Answer Sheet	88	12.73	4.30		
Language Expression					
Consumable Bklt.	98	20.32	4.82	.48	.65
Answer Sheet	88	19.84	5.09		
Total Language					
Consumable Bklt.	95	64.56	12.32	.79	.40
Answer Sheet	86	63.77	14.10		

TABLE 4B: COMPARISON OF MEAN NUMBER OF BLANK RESPONSES OF MEDIUM ABILITY THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Blank Responses	S.D.	Mean Difference	*t*
Reading Vocabulary					
Consumable Bklt.	96	.83	2.18	-1.16	-2.24 *
Answer Sheet	88	1.99	4.50		
Reading Comprehension					
Consumable Bklt.	96	.98	2.14	-.88	-1.70
Answer Sheet	88	1.86	4.58		
Spelling					
Consumable Bklt.	96	.55	1.56	-2.08	-3.91
Answer Sheet	88	2.63	4.93		
Language Mechanics					
Consumable Bklt.	97	.23	.70	.12	1.24
Answer Sheet	88	.11	.51		
Language Expression					
Consumable Bklt.	98	.34	.69	.01	.06
Answer Sheet	88	.33	.92		

* Asterisk indicates that mean difference is significantly different from zero at the .05 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

TABLE 5A: COMPARISON OF MEAN RAW SCORES OF HIGH ABILITY THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Raw Score	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	28	34.68	3.30	.27	.28
Answer Sheet	27	34.41	3.78		
Reading Comprehension					
Consumable Bklt.	28	39.79	3.61	.68	.64
Answer Sheet	27	39.11	4.22		
Total Reading					
Consumable Bklt.	28	74.46	5.96	.94	.54
Answer Sheet	27	73.52	7.09		
Spelling					
Consumable Bklt.	28	40.82	2.89	1.52	1.58
Answer Sheet	27	39.30	4.18		
Language Mechanics					
Consumable Bklt.	27	17.44	2.10	.25	.39
Answer Sheet	26	17.19	2.56		
Language Expression					
Consumable Bklt.	27	27.19	2.11	.69	1.09
Answer Sheet	26	26.50	2.45		
Total Language					
Consumable Bklt.	27	85.41	5.72	2.53	1.47
Answer Sheet	26	82.88	6.75		

TABLE 5B: COMPARISON OF MEAN BLANK RESPONSES OF HIGH ABILITY THIRD GRADERS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Blank Responses	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	28	.07	.26	-.08	-.77
Answer Sheet	27	.15	.46		
Reading Comprehension					
Consumable Bklt.	28	.18	.39	.14	1.70
Answer Sheet	27	.04	.19		
Spelling					
Consumable Bklt.	28	.29	.66	.18	1.24
Answer Sheet	27	.11	.32		
Language Mechanics					
Consumable Bklt.	27	.04	.19	-.38	-1.02
Answer Sheet	26	.42	1.96		
Language Expression					
Consumable Bklt.	27	.11	.32	-.08	-.72
Answer Sheet	26	.19	.49		

* Asterisk indicates that mean difference is significantly different from zero at the .05 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

There were no significant differences between mean raw scores of students using the consumable booklets and those using separate answer sheets among any of the three ability groupings. Students classified as low ability students seem to score about as well on one test format as the other. This is true even though low ability students using consumable booklets scored slightly higher in some cases than did those using separate answer sheets and the reverse occurred in several other cases. These differences were considered statistically insignificant. They can be considered a function of chance variation in test scores.

There were only three instances in the analyses of blank responses where significant differences were observed. These analyses are summarized in Tables 3B, 4B, and 5B. A significant difference among low ability students using the consumable booklets and those using separate answer sheets occurred in the reading comprehension subtest (Table 3B). Those using the separate answer sheet left a significantly higher number of blank responses, indicating a slower work rate, on the reading comprehension subtest. A similar finding was observed among medium ability students on the reading vocabulary and spelling subtests (Table 4B). Students using separate answer sheets were found to have left a significantly higher number of blank responses than those using consumable booklets. As can be seen by observing Table 5B, there were no statistically significant differences among high ability students in relation to the number of blank responses on the two test formats.

A sixth analysis was undertaken which included only the third grade EMH's, LD's, and retentions. In contrast to the second analysis previously discussed

which eliminated all students who have had some identified learning problem, this analysis includes only those who have been identified as having some kind of a learning problem. As previously discussed, the emphasis of the testing related portions of the Educational Improvement Act is on identifying educational needs. These students are the ones most likely to be identified as having high priority educational needs, along with those students of low ability. (These two groups overlap; they include many of the same students.) These are the students on which attention will be focused. The results of this analysis are summarized in Tables 6A and 6B.

One significant difference in mean raw scores was observed as can be seen in Table 6A. Students with identified learning problems who used the consumable booklet format scored significantly higher on the spelling subtest than did students using the separate answer sheet. This was the only instance in which a significant difference in mean raw test scores was observed. There have been several cases in which significant differences in work rates (mean number of blank responses) were observed. It is interesting to note that in Table 6B there are no significant differences in the difference between mean number of blank responses. Thus, the work rates of students having identified learning problems did not differ significantly on the two testing formats.

The upper limits of each decile was also calculated and reported in Table 7. This was done for descriptive purposes only. Calculations were performed only on the total reading and total language scores. As can be observed in Table 7, the differences between these upper limits would appear to be somewhat greater at the lower deciles than at the upper deciles. These kinds of results should not be considered too alarming unless one intends to test some students with consumable booklets and some with separate answer sheets. If comparisons among students using the two test formats were to be made, one might give some

TABLE 6A: COMPARISON OF MEAN RAW SCORES OF THIRD GRADE EMH'S, LD'S, AND RETENTIONS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Raw Score	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	24	18.50	8.31	3.05	1.26
Answer Sheet	22	15.45	8.08		
Reading Comprehension					
Consumable Bklt.	24	18.54	9.68	1.27	.44
Answer Sheet	22	17.27	9.97		
Total Reading					
Consumable Bklt.	24	37.04	17.42	-.69	.84
Answer Sheet	22	32.73	17.44		
Spelling					
Consumable Bklt.	23	29.26	7.18	4.76	2.13*
Answer Sheet	22	24.50	7.84		
Language Mechanics					
Consumable Bklt.	23	8.70	5.13	.93	.63
Answer Sheet	22	7.77	4.62		
Language Expression					
Consumable Bklt.	23	14.78	6.40	1.46	.82
Answer Sheet	22	13.32	5.47		
Total Language					
Consumable Bklt.	22	53.27	15.92	7.68	1.66
Answer Sheet	22	45.59	14.71		

TABLE 6B: COMPARISON OF MEAN NUMBER OF BLANK RESPONSES OF THIRD GRADE EMH'S, LD'S, AND RETENTIONS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Subtest	N	Mean Blank Responses	S.D.	Mean Difference	t
Reading Vocabulary					
Consumable Bklt.	24	1.54	3.35	-2.05	-1.34
Answer Sheet	22	3.59	6.60		
Reading Comprehension					
Consumable Bklt.	24	3.13	4.79	.95	.68
Answer Sheet	22	2.18	4.64		
Spelling					
Consumable Bklt.	23	1.26	2.14	-2.19	-1.55
Answer Sheet	22	3.45	6.43		
Language Mechanics					
Consumable Bklt.	23	.48	.99	.43	2.00
Answer Sheet	22	.05	.21		
Language Expression					
Consumable Bklt.	23	1.78	2.88	1.19	1.80
Answer Sheet	22	.59	1.18		

* Asterisk indicates that mean difference is significantly different from zero at the .05 level. If no asterisk appears, then the mean difference is considered statistically insignificant.

TABLE 7: COMPARISON OF THE UPPER LIMITS OF THIRD GRADE RAW SCORE DECILES BETWEEN STUDENTS USING CONSUMABLE BOOKLETS AND THOSE USING SEPARATE ANSWER SHEETS

Decile	Reading		Language	
	C.B.	S.A.S.	C.B.	S.A.S.
1	23	22	42	37
2	33	30	50	46
3	40	37	58	53
4	48	45	61	60
5	55	53	66	65
6	61	59	72	72
7	65	65	77	76
8	72	72	81	80
9	75	76	88	84
10	82	83	92	92

C.B. = Consumable Booklet
S.A.S.= Separate Answer Sheet

consideration to developing two sets of norms. This was a recommendation of Aleyideino. However, so long as all students use the same test format, this is not a major concern.

In summary, there is no evidence as collected for this study which would justify the use of one test format over the other at the third grade. Either test format would seem appropriate. Students with identified learning problems did experience greater difficulty with the spelling subtest on the separate answer sheet. This was the only noticeable difference in test scores. In several instances, significant differences in work rates were observed. As indicated by blank responses, students worked faster using the consumable booklet. However, these differences did not seem to cause significant differences in the mean raw scores of students using the two different test formats.

TEACHER SURVEY

Teachers who administered the tests were asked several questions after testing was completed. When asked if they thought that most third graders could handle the separate answer sheets, a majority of those teachers administering the separate answer sheet format of the test responded affirmatively. When asked if they thought those students who performed poorly using the separate answer sheet would have done better using the consumable booklet, teacher responses were mixed. Some thought that test format would make no difference. Other teachers thought that some students who performed poorly on the separate answer sheet might have benefitted from the use of the consumable booklet, especially those students classified as LD's.

SUMMARY AND CONCLUSIONS

When the scores of third grade students used in this sample were analyzed all together, with no delineation whatsoever, there were no significant differences between students marking their responses on consumable booklets and those marking their responses on separate answer sheets. In all CTBS subtests except one, language mechanics, students using consumable booklets scored slightly higher than those using separate answer sheets. In language mechanics, the students using the separate answer sheets scored slightly higher. However, in no case was the mean raw score difference significant. In reading vocabulary and spelling subtests, students using the answer sheet left a significantly higher number of blank responses than did those using the consumable booklets. The same significant findings were observed when the EMH's, LD's and retentions were excluded.

When only low ability students (those scoring one standard deviation or more below the mean on verbal IQ) were analyzed, there were no significant differences in mean scores between those using the consumable booklet and those using the separate answer sheet. A significant difference among low ability students using the consumable booklets and those using separate answer sheets as to the number of blank responses was observed in the reading comprehension subtest. Those using the separate answer sheet left a significantly high number of blank responses, indicating a slower work rate, on the reading comprehension subtest.

A similar finding was observed among medium ability students (those scoring within one standard deviation above or below the mean on the verbal IQ) on the reading vocabulary and spelling subtests. Students using separate answer sheets were found to have left a significantly higher number of blank responses than those using consumable booklets, on these two subtests.

This significant difference in the number of blank responses did not appear on any of the subtests when the scores of high ability students (those scoring one standard deviation or more above the mean on the verbal IQ) were analyzed.

In analyzing only those students having some identified learning problem, EMH, LD, or retention, a significant difference in mean raw scores was observed on one subtest only. Students with identified learning problems who used the consumable booklet format scored significantly higher on the spelling subtest than did students using the separate answer sheet. The work rates of students having identified learning problems, as indicated by the number of blank responses, did not differ significantly on the two testing formats.

The research performed here indicates there is no evidence to justify the use of one format over the other at the third grade level. Either the use of the consumable booklet or the non-consumable booklet with the separate answer sheet would be appropriate. The results of this research along with the fact that the teachers involved in this project feel positive about the use of the separate answer sheet leaves little doubt that the State Department of Education could use this more economically feasible format without any detrimental affects.

REFERENCES

Aleyideino, Samuel C., "The Effects of Response Methods and Item Types Upon Working Time Scores and Grade Equivalent Scores of Pupils Differing in Levels of Achievement." Unpublished doctoral dissertation, the University of Iowa, June, 1968.

Barr, Anthony J. and others, A User's Guide to SAS 76, (Raleigh, N.C.: Sparks Press, 1976), pp. 275-277.

Cresap, McCormick and Paget, "Report to Kentucky Commission on Public Education." Study submitted in 1964.

Hieronymus, A.M., "Research on the Use of a Separate Answer Sheet in Grade 3." Paper presented for the Test Department of Houghton Mifflin Co., Iowa City, Iowa. (Undated)

Printed with State funds